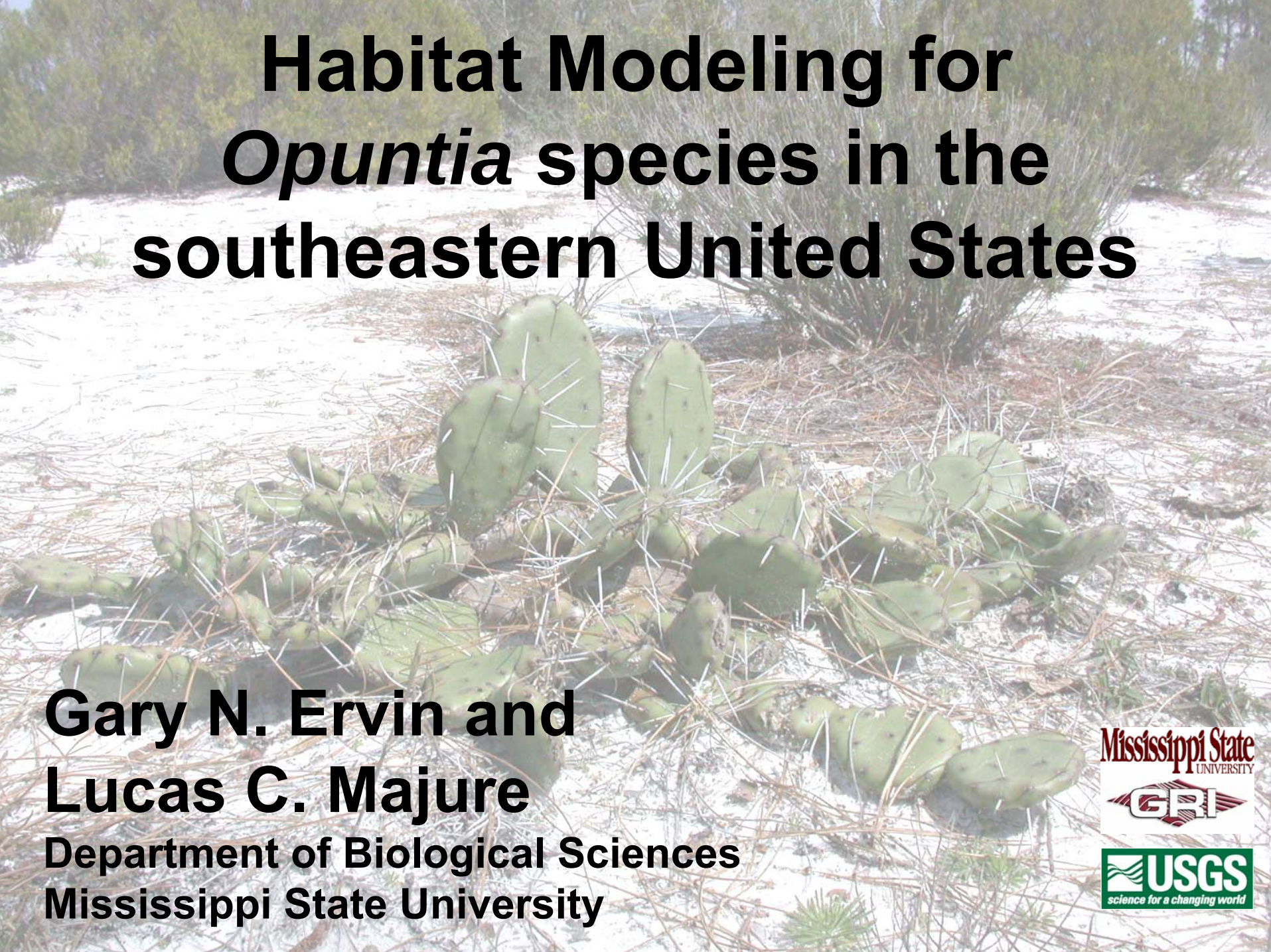
Habitat Modeling for *Opuntia* species in the southeastern United States



Gary N. Ervin and Lucas C. Majure
Department of Biological Sciences
Mississippi State University

Mississippi State University
GRI
USGS science for a changing world

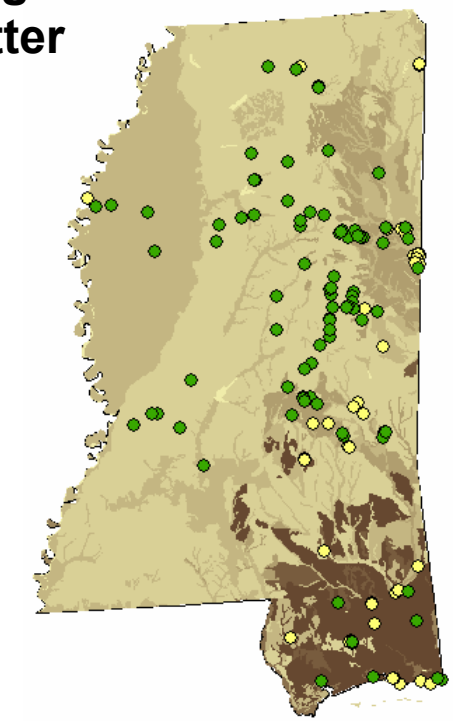
Rationale

Accurate predictive models for *Opuntia* habitat will facilitate efforts at locating and monitoring the progress of *Cactoblastis* invasion.

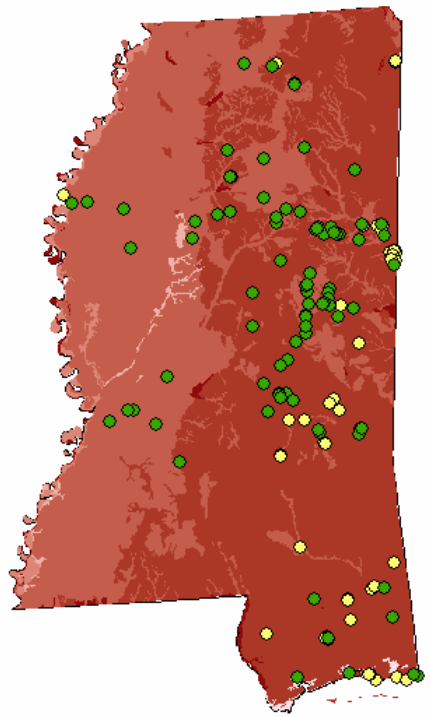
Objective

Use geospatial data layers to predict cactus presence via logistic regression and GIS

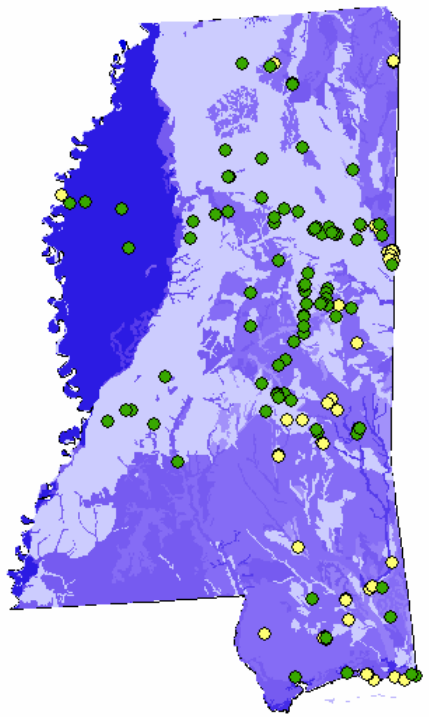
Maximum soil organic matter



Minimum pH



Maximum CEC



QUESTIONS

1. How to select the best model?

   Model Fit *versus*

   Model Adequacy

2. What spatial extent is most appropriate?

   Local *versus*

   State or Regional models

MODEL SELECTION

Model Fit

Does the model fit the available data?

- usually based on the data used to derive the model, e.g., likelihood tests, AIC, BIC

Model Accuracy

Does the model adequately depict reality?

- based on accuracy assessment criteria, ideally with an independent data set

MODEL FIT:

Information-theoretic analyses, e.g., Akaike Information Criterion (AIC)

$$AIC_c = -2 \times \left(-\frac{n}{2} \log\left(\frac{RSS}{n-(p+1)} \right) \right) + 2K + \left(\frac{2K(K+1)}{n-K-1} \right)$$

$$\Delta AIC_c = AIC_{ci} - AIC_{cBest}$$

Support for Model i:

ΔAIC = 0 – 2 : Substantial

ΔAIC = 4 – 7 : Considerably less

ΔAIC > 10 : Essentially none

$$w_i = \frac{\exp\left(-\frac{\Delta_i}{2} \right)}{\sum_{r=1}^{R} \exp\left(-\frac{\Delta_r}{2} \right)}$$

(Burnham & Anderson 2002)

MODEL FIT:
Information-theoretic analyses, e.g., Akaike Information Criterion (AIC)

$$AIC_c = -2 \times \left(-\frac{n}{2} \log \left(\frac{RSS}{n-(p+1)} \right) \right) + 2K + \left(\frac{2K(K+1)}{n-K-1} \right)$$

$$\Delta AIC_c = AIC_{ci} - AIC_{cBest}$$

Support for Model i:

ΔAIC = 0 – 2 : Substantial

ΔAIC = 4 – 7 : Considerably less

ΔAIC > 10 : Essentially none

$$w_i = \frac{\exp\left(-\frac{\Delta_i}{2}\right)}{\sum_{r=1}^{R} \exp\left(-\frac{\Delta_r}{2}\right)}$$

(Burnham & Anderson 2002)

MODEL FIT:
Information-theoretic analyses, e.g., Akaike Information Criterion (AIC)

$$AIC_c = -2 \times \left(-\frac{n}{2} \log\left(\frac{RSS}{n-(p+1)} \right) \right) + 2K + \left(\frac{2K(K+1)}{n-K-1} \right)$$

$$\Delta AIC_c = AIC_{ci} - AIC_{cBest}$$

Support for Model i:

$\Delta AIC = 0 - 2$: Substantial

$\Delta AIC = 4 - 7$: Considerably less

$\Delta AIC > 10$: Essentially none

(Burnham & Anderson 2002)

$$w_i = \frac{\exp\left(-\frac{\Delta_i}{2} \right)}{\sum_{r=1}^{R} \exp\left(-\frac{\Delta_r}{2} \right)}$$

MODEL FIT:
Information-theoretic analyses, e.g., Akaike Information Criterion (AIC)

$$AIC_c = -2 \times \left(-\frac{n}{2} \log\left(\frac{RSS}{n-(p+1)} \right) \right) + 2K + \left(\frac{2K(K+1)}{n-K-1} \right)$$

$$\Delta AIC_c = AIC_{ci} - AIC_{cBest}$$

$$w_i = \frac{\exp\left(-\frac{\Delta_i}{2} \right)}{\sum_{r=1}^{R} \exp\left(-\frac{\Delta_r}{2} \right)}$$

Support for Model i:

$\Delta AIC = 0 - 2$: Substantial

$\Delta AIC = 4 - 7$: Considerably less

$\Delta AIC > 10$: Essentially none

(Burnham & Anderson 2002)

MODEL ACCURACY:

Overall success	**influenced by prevalence and assigns high accuracy to rare species**
Sensitivity and Specificity	**independent of prevalence but not adjusted for chance**
Cohen's kappa	**accounts for chance and omission/commission errors but influenced by prevalence**
True Skill Statistic	**similar to kappa but thought to be independent of prevalence**

The last two range from -1 to +1; +1 = perfect agreement.

A SPECIFIC EXAMPLE

Data

***Opuntia* presence-absence (189 points):**

- ***O. humifusa* – 99 presence & 90 absence (somewhat balanced)**

Soil (STATSGO geospatial data layers):

- **available water content, bulk density, CEC, clay content, organic matter content, permeability, pH (minimum & maximum per MU)**

Analyses

Correlation analyses among soil parameters

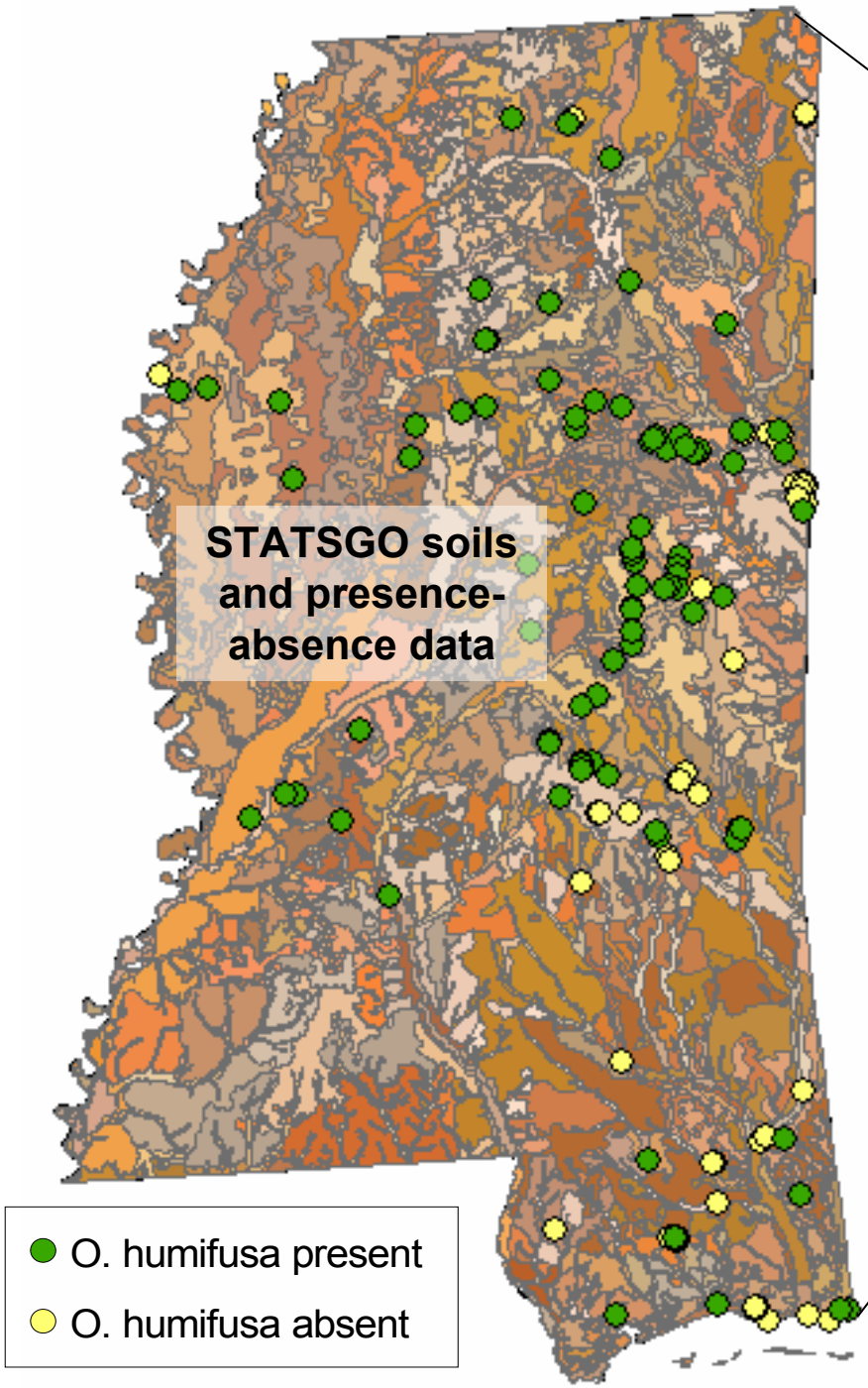
- → **resulted in 46 logistic regression models**

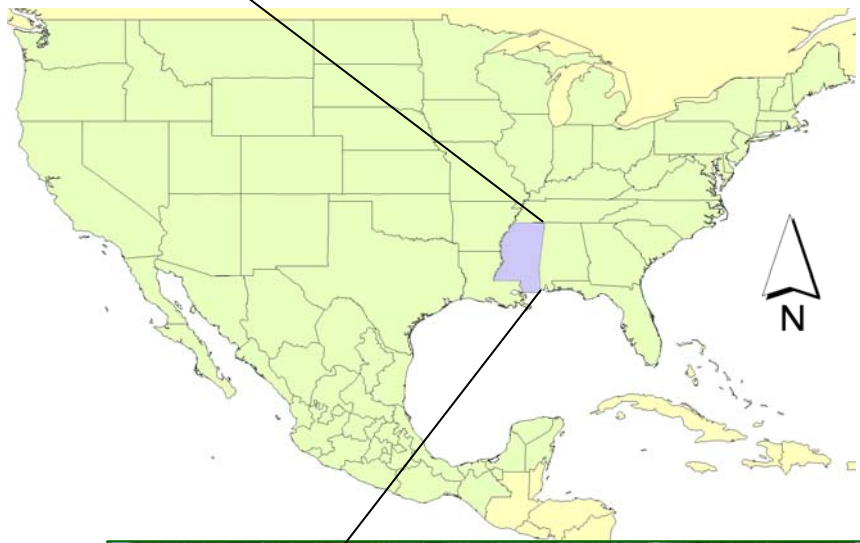
AIC analyses to compare resulting models

Calculated all five model accuracy criteria

- **(will focus here on kappa & TSS)**

Opuntia humifusa





Relative importance of soil parameters

AIC approach

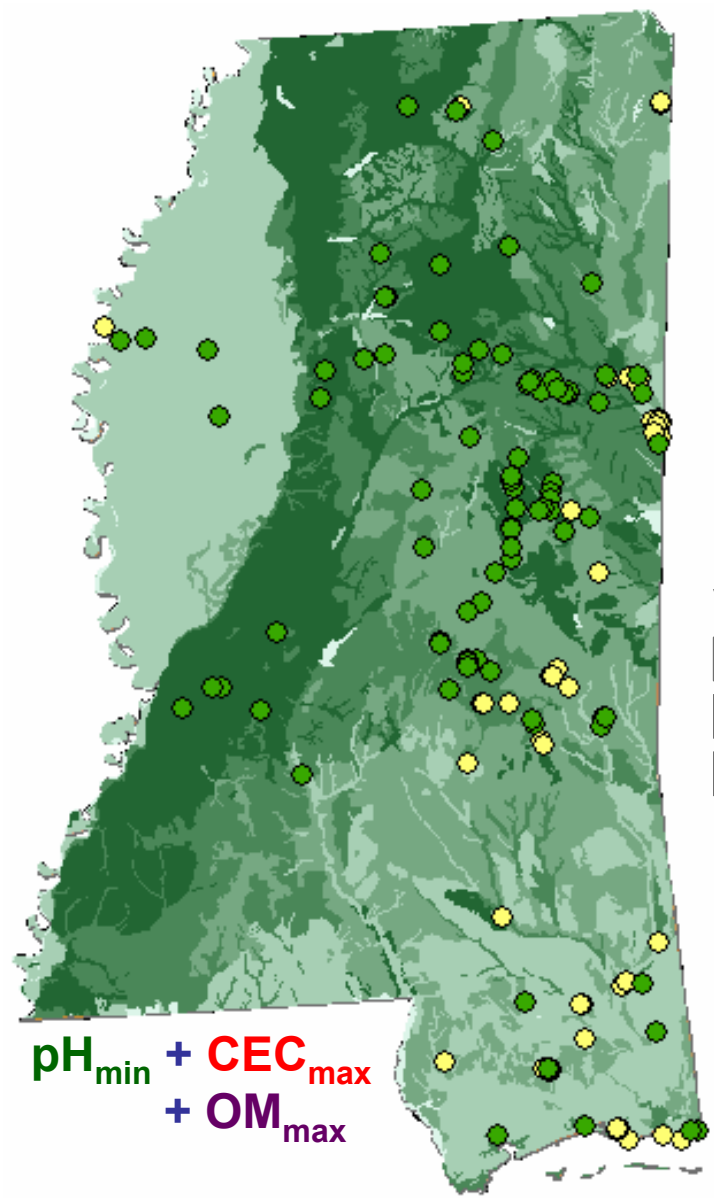
Soil parameters in model	ΔAICc	Akaike weight (w)
pH_{min} + CEC_{max} + OM_{max}	0.0	0.20
$Perm_{max}$	0.1	0.19
pH_{min} + CEC_{max} + OM_{max} + $Perm_{min}$	0.6	0.15
pH_{min} + CEC_{max} + AWC_{max}	1.2	0.11
pH_{min} + CEC_{max} + $Perm_{min}$	1.4	0.10
pH_{min} + CEC_{max}	1.6	0.09

Relative importance of soil parameters

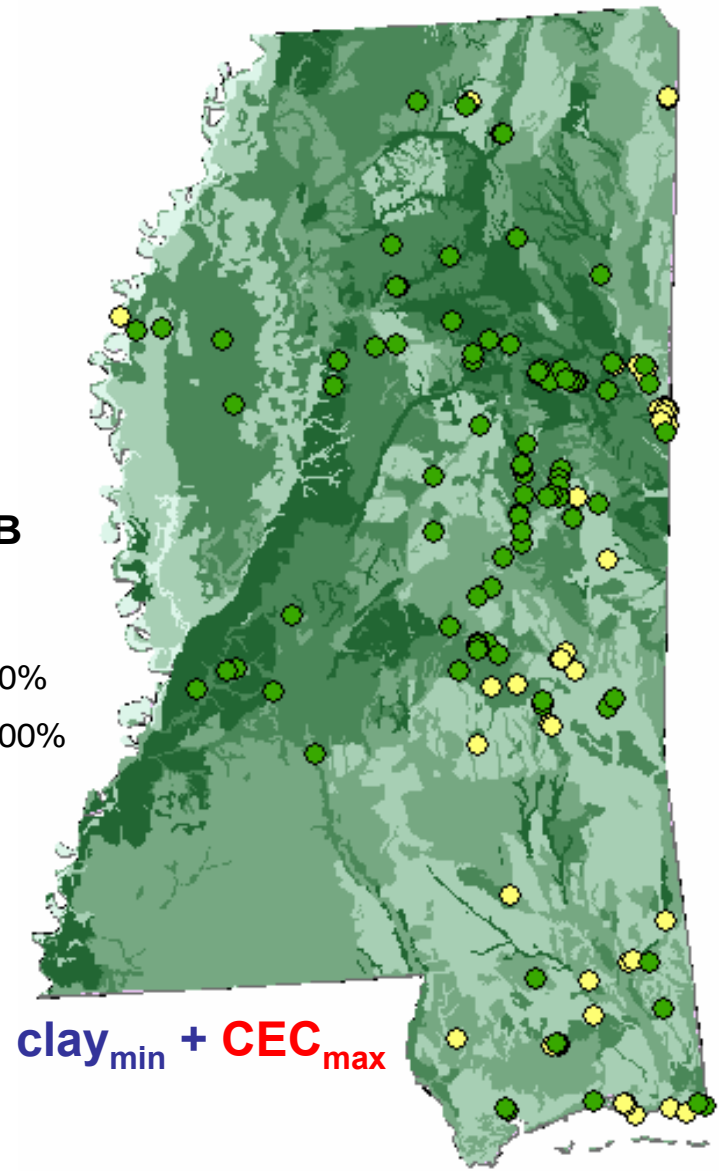
Accuracy criteria

	Soil parameters in model	kappa	TSS
	$clay_{min}$ + CEC_{max}	0.41	0.42
	$clay_{min}$ + CEC_{max} + OM_{min}	0.40	0.40
2 →	$Perm_{max}$	0.38	0.38
	$Perm_{max}$ + OM_{min}	0.38	0.38
	pH_{max} + CEC_{max}	0.37	0.37
1 →	pH_{min} + CEC_{max} + OM_{max}	0.37	0.37
4 →	pH_{min} + CEC_{max} + AWC_{max}	0.37	0.37

Model fit



Model Accuracy



O. humifusa PROB

Value

- 0 to 20%
- 20 to 40%
- 40 to 60%
- 60 to 80%
- 80 to 100%

SPATIAL EXTENT

Available data

- **Data may be restricted in distribution or**
- **Presences and absences may be inadequately dispersed**

Objectives

- **Targeting specific areas for surveys versus**
- **Estimating potential distributions**

A SECOND EXAMPLE

Data

***Opuntia* presence-absence (33 points):**

- ***O.* affinis *grandiflora* – 14 presence & 19 absence (again somewhat balanced)**

Soil (STATSGO geospatial data layers):

- **available water content, bulk density, CEC, clay content, organic matter content, permeability, pH (minimum & maximum per MU)**

Analyses

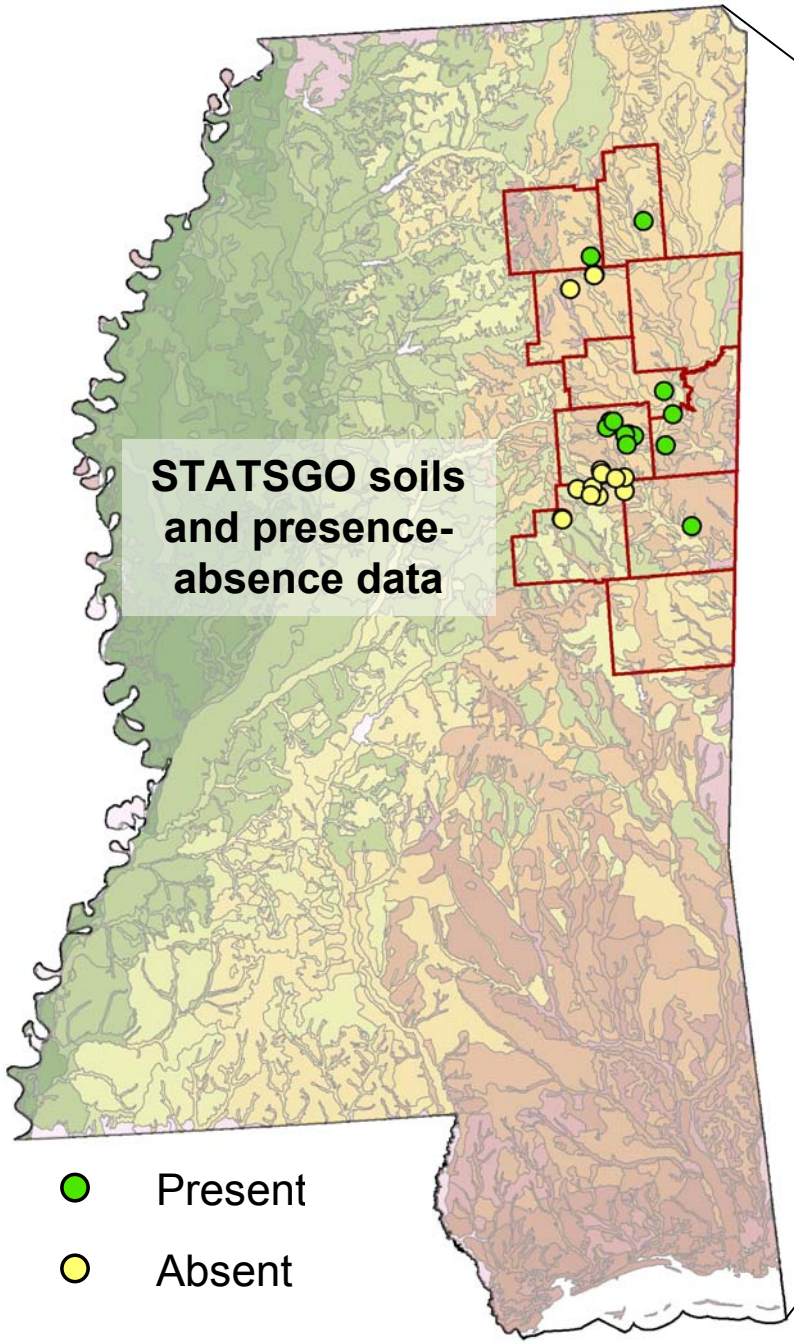
Correlation analyses among soil parameters and with *Opuntia* presence

- **→ resulted in 19 logistic regression models**

AIC analyses to compare resulting models

Calculated five model accuracy criteria

Opuntia aff. *grandiflora*



Relative importance of soil parameters

AIC approach

Soil parameters in model	ΔAICc	Akaike weight (w)
$BulkDens_{min}$ + pH_{max}	0.0	0.49
$BulkDens_{min}$ + $Clay_{max}$	1.2	0.27
$BulkDens_{min}$ + pH_{max} + pH_{min}	2.8	0.12
$BulkDens_{min}$ + $Clay_{max}$ + pH_{min}	3.9	0.07

Relative importance of soil parameters

Accuracy criteria

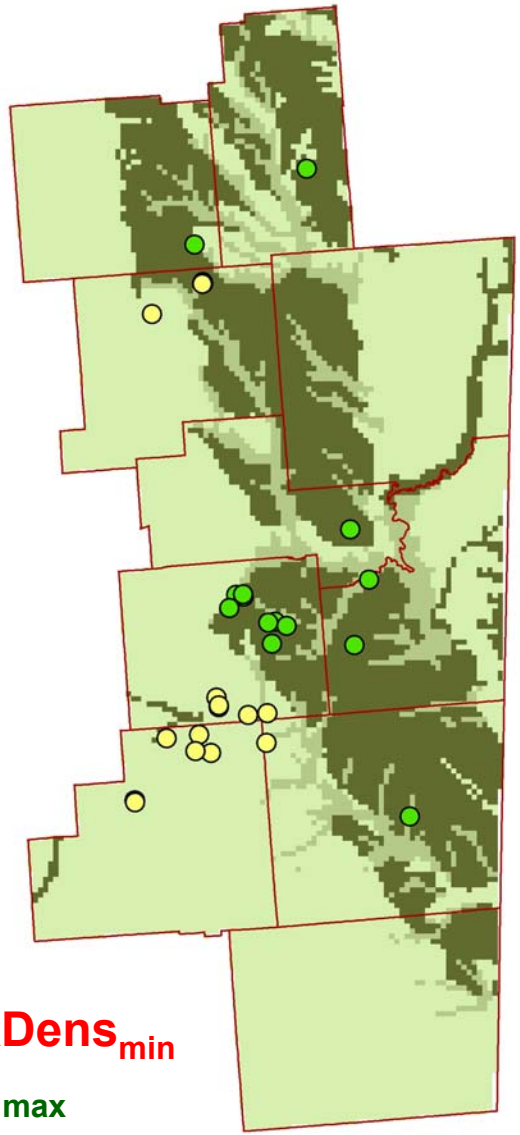
Soil parameters in model	kappa	TSS
$BulkDens_{min}$ + pH_{max}	0.88	0.88
$BulkDens_{min}$ + $Clay_{max}$	0.88	0.88
$BulkDens_{min}$ + pH_{max} + pH_{min}	0.88	0.88
$BulkDens_{min}$ + $Clay_{max}$ + pH_{min}	0.88	0.88

Success = 94%

Sensitivity = 93% and Specificity = 95%

(versus 70%, 63%, and 78% for best *O. humifusa* models)

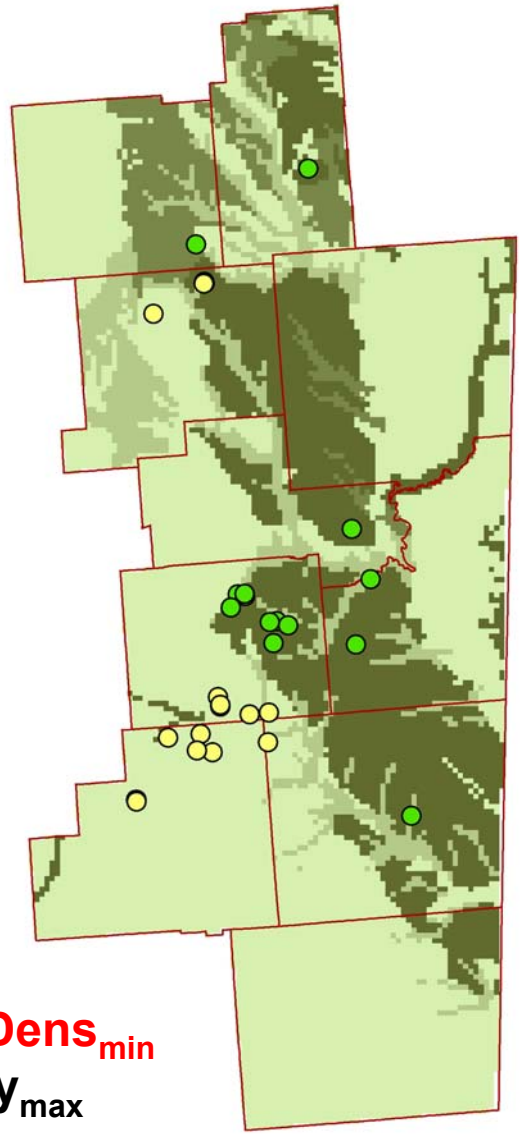
Top two models based on AIC



BulkDens$_{min}$
+ pH$_{max}$

Presence Prob
Value

	0.0 - 0.2
	0.2 - 0.4
	0.4 - 0.6
	0.6 - 0.8
	0.8 - 1



BulkDens$_{min}$
+ Clay$_{max}$

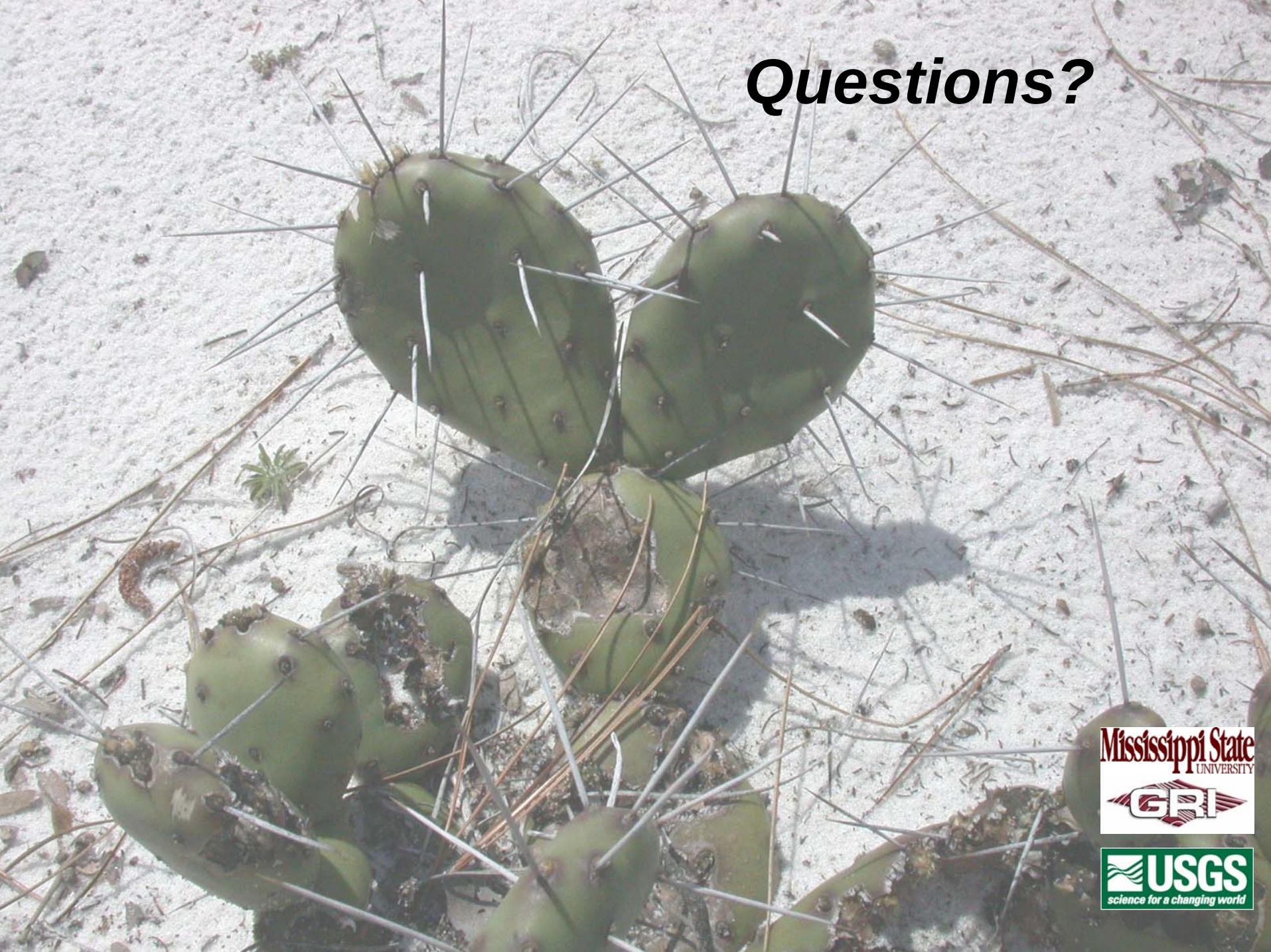
SUMMARY

Model selection

- Some degree of agreement between model fit and model accuracy criteria

Spatial extent

- Models using the smaller spatial extent exhibited greater agreement between fit and accuracy criteria
- Models using the smaller extent resulted in higher values for assessment criteria


Questions?
Mississippi State
UNIVERSITY
GRI
USGS
science for a changing world